

# Performance Bounds for Grouped Incoherent Measurements in Compressive Sensing

Adam C. Polak, Marco F. Duarte, and Dennis L. Goeckel\*

March 3, 2013

## Abstract

Compressive sensing (CS) allows for acquisition of sparse signals at sampling rates significantly lower than the Nyquist rate required for bandlimited signals. Recovery guarantees for CS are generally derived based on the assumption that measurement projections are selected independently at random. However, for many practical signal acquisition applications, this assumption is violated as the projections must be taken in groups. In this paper, we consider such applications and derive requirements on the number of measurements needed for successful recovery of signals when groups of dependent projections are taken at random. We find a penalty factor on the number of required measurements with respect to the standard CS scheme that employs conventional independent measurement selection and verify the predicted penalty through simulations.

## 1 Introduction

Modern receivers and sensors need to process an enormous amount of bandwidth to satisfy continuously growing demands on communication and sensing systems. Due to the complexity and high power consumption of hardware at large bandwidths, a number of innovative approaches for signal acquisition have recently emerged, including a class based on compressive sensing (CS). In CS approaches, the full signal bandwidth is not converted, hence avoiding the costly hardware; rather, prior knowledge of a concise signal model allows the recovery to focus only on signal aspects relevant to feature extraction. In particular, if there exists a basis in which a signal of interest can be represented sparsely (i.e., it can be fully characterized with a small number of coefficients), then it is possible to obtain all information needed for successful reconstruction of the

---

\*This paper is based in part upon work supported by the National Science Foundation under grants CNS-0905349 and ECCS-1201835. We thank Waheed Bajwa, Mark Rudelson, and Joel Tropp for helpful comments.

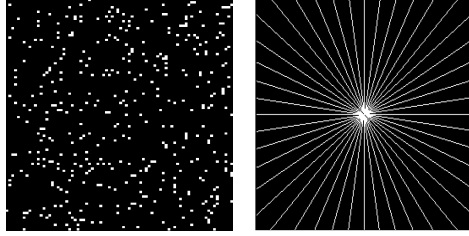


Figure 1: *Left: Independently random 2-D sampling. Right: Radial acquisition trajectories used for MRI, which group measurement selections into slices of the 2-D Fourier domain.*

signal from a relatively small number of randomized incoherent measurements [1]. This number is often much smaller than the number of samples implied by the Nyquist sampling rate for representation of all bandlimited signals.

Most CS contributions assume independent randomness in the measurement projections that is exploited to derive bounds on the number of projections needed for successful recovery. However, for many practical signal acquisition applications, this assumption is violated as the projection measurements must be selected in groups. As an example, consider Magnetic Resonance Imaging (MRI), where the measurements in the 2-D Fourier space cannot be taken at random but need to follow sampling trajectories that satisfy hardware and physiological constraints: the radial acquisition trajectories of MRI shown in Fig. 1 are known to be especially suitable for high-contrast objects. Using such sampling trajectories clearly introduces structure into the measurement process and hence violates a key assumption underlying the standard analysis of CS schemes.

In this work, we derive bounds on the number of measurements needed for successful recovery of signals when the random projection measurements are structured into predefined groups. We introduce a metric that upper bounds the multiplicative penalty on the number of required measurements introduced by grouping with respect to conventional CS acquisition employing independently random measurement selection; such a metric is dependent on the sparse signal support and might be useful in the design of many practical signal acquisition systems with grouped measurement structures. While this metric cannot currently be evaluated in a closed form, we employ a computationally feasible method that provides lower and upper bounds on its value. We also verify via simulations the penalty predicted by the proposed metric.

## 2 Background

### 2.1 Compressive Sensing

Consider the acquisition of an  $N \times 1$  signal vector  $\underline{x}$ . Assume that  $\underline{x}$  is known to be sparse in some basis; that is, we say the signal  $\underline{x}$  is  $K$ -sparse for some integer  $K$  if  $\underline{x}$  has a representation  $\underline{c} = U^H \underline{x}$  having only  $K$  non-zero entries in some known orthonormal basis  $U$ , although the value and location of those non-zero entries may be unknown. In the CS framework, we acquire the  $M \times 1$  output  $y = \Phi \underline{x}$ , for some  $M \ll N$ , where  $\Phi$  is the measurement matrix. According to CS theory, given certain constraints on  $\Phi$  and  $M$ ,  $\underline{x}$  can be reconstructed from  $y$  with high probability.

### 2.2 Incoherent Measurements

Given an orthonormal measurement basis  $V$ , a  $K$  sparse signal  $\underline{x} = U \underline{c}$ , sparse in some known orthonormal basis  $U$  can be reconstructed successfully from a set of  $M$  independently drawn random samples  $\Omega \subseteq \{1, \dots, N\}$  of  $y = V^H U \underline{c}$  with probability not lower than  $1 - \delta$ , for any  $\delta > 0$ , as long as the number of samples is large enough. Define  $A = V^H U$  and denote by  $A_\Omega$  the matrix built from the  $M$  rows of  $A$  corresponding to the index set  $\Omega$ . Define the coherence  $\mu(A)$  of the matrix  $A$  as  $\mu(A) = \max_{i,j} |A(i, j)|$ , which has range  $\mu(A) \in [\frac{1}{\sqrt{N}}, 1]$  [1]. A pair of bases  $V$  and  $U$  for which the minimal value of  $\mu(A)$  is achieved is referred to as a *perfectly incoherent* pair of bases.

When the elements of  $\Omega$  are drawn independently at random, it can be shown that the number  $M$  of measurements required for successful recovery of sparse  $\underline{x}$  depends on the coherence of the matrix  $A$ .

**Theorem 1.** [1] *Let  $A$  be an  $N \times N$  orthogonal matrix ( $A^H A = I$ ) with coherence  $\mu(A)$ . Fix an arbitrary subset  $T$  of the signal domain. Choose a subset  $\Omega$  of the measurement domain of size  $|\Omega| = M$  and a sign sequence  $z$  on  $T$ , both uniformly at random over all possible choices. Suppose that*

$$M \geq \text{Const} \cdot N \mu^2(A) |T| \log(N/\delta). \quad (1)$$

*Then with probability exceeding  $1 - \delta$ , every signal  $\underline{c}_0$  supported on  $T$  with signs matching  $z$  can be recovered from  $y = A_\Omega \underline{c}_0$  by solving the linear program*

$$\min_{\underline{c}} \|\underline{c}\|_1 \quad s.t. \quad A_\Omega \underline{c} = A_\Omega \underline{c}_0. \quad (2)$$

Theorem 1 shows that the number of measurements required for successful recovery of a sparse signal scales linearly with the signal's sparsity, but only logarithmically with its length, as long as  $V$  and  $U$  are perfectly incoherent.

### 2.3 Grouped Incoherent Measurements

In certain applications, the assumptions of Theorem 1 are violated as measurements must be taken in groups instead of independently at random. More

specifically, divide the set of  $N$  rows of the measurement basis  $V$  into  $N/G$  disjoint groups  $\mathcal{G}_i$ ,  $i = 1, \dots, N/G$ , of size  $G$  each. Note that it will still be possible to take a set of measurements  $\Omega$  for a signal, following Theorem 1, by selecting  $M/G$  groups out of the  $N/G$  groups available, independently at random.<sup>1</sup> We say that such a process provides a *grouped incoherent measurement scheme*. Grouped incoherent measurement schemes can be seen as a generalization of the standard incoherent measurement scheme used in Theorem 1 by setting  $G = 1$ .

## 2.4 Example Applications

*Interference-Robust Compressive Wideband Receiver:* One important example application for a grouped incoherent measurement scheme is an interference-robust compressive wideband receiver.<sup>2</sup> If a large communication bandwidth is employed, interference is nearly always present. More importantly, it is common for the signal of interest to be buried in an interferer that is orders of magnitude stronger. This might force the receiver’s RF front end into the nonlinear range and cause intermodulation distortion that makes the interference cancellation methods based on interference null space projection [2] ineffective. As an alternative, we may opt to perform sampling only at times in which the RF front end is not saturated and exhibits linear behavior, e.g., at times when the interferer’s value is small. A typical interferer is modulated; therefore, while its first few zero-crossings can be considered as random, the remaining set of subsequent zero-crossings are dictated by the frequency of the interferer’s carrier. Therefore, a sampling approach that aims to operate within the linear region of the RF front end results in a grouped incoherent measurement scheme, in effect providing an interference-robust compressive wideband receiver.

*Medical Imaging:* There are a multitude of medical imaging applications that rely on tomography principles, where CS can be applied to reduce the number of measurements required for accurate image recovery [3]; common examples include MRI and computed axial tomography (CAT). In tomographic imaging, the 2-D image measurements obtained via a tomographic scan correspond to samples of the Radon transform of the image. These samples can be grouped by orientation and processed in groups via the discrete Fourier transform. According to the projection slice theorem, the output of this transformation provides samples of the image’s 2-D discrete Fourier transform along a line running through the origin (cf. Fig. 1). Thus, the measurements obtained correspond to a grouped measurement in the 2-D Fourier transform domain of the image, and groups can be selected independently by selecting tomographic scan orientations independently.

*Multi-dimensional signals and signal ensembles:* For signals spanning many physical dimensions, such as space, time, spectrum, etc., it is often difficult to design CS acquisition devices that can calculate random projections involving

---

<sup>1</sup>We assume that  $G$  divides  $M$  and  $N$  for simplicity.

<sup>2</sup>We thank Robert Jackson for suggesting this architecture for addressing the interference problem in wideband receivers.

all signal samples. Instead, it is commonly easier to modify the CS acquisition process so that it is applied separately to each piece of a partition of the multidimensional signal. Examples include hyperspectral imaging and video acquisition, sensor networks, and synthetic aperture radar [4, 5]. Consider in particular the case where the choice of measurements used for each partition comes from a single orthonormal basis and is shared among partitions, introducing structure in the measurements. For example, a compressive video camera may use the same incoherent projections on each frames in the video sequence. The resulting global measurement basis is downsampled in a group-structured fashion. The grouped incoherent measurement framework can be applied when a single orthonormal basis is used for compression of the entire multidimensional signal [4].

### 3 Performance Analysis for Grouped Incoherent Measurements

#### 3.1 Performance Metric

The grouped incoherent measurement scheme introduced in Section 2.3 violates the assumptions of Theorem 1 and causes an increase of the number of measurements needed for successful recovery of sparse signals. Such a penalty factor depends on the structure of the groups  $\mathcal{G} = \{\mathcal{G}_1, \dots, \mathcal{G}_{N/G}\}$ , on the product of the measurement and transformation basis  $A = V^H U$ , and on the set  $T$  defining the sparse signal support. We define a penalty factor

$$\gamma(A, T, \mathcal{G}) = \max_{i \in 1, \dots, N/G} \|\overline{A_{\mathcal{G}_i T}}\|_{2 \rightarrow 1}, \quad (3)$$

where  $\|M\|_{p \rightarrow q} = \max_f \|Mf\|_q / \|f\|_p$  denotes the  $p \rightarrow q$  operator norm of the matrix  $M$ ,  $\overline{M}$  denotes the matrix  $M$  after row normalization, and  $A_{\mathcal{G}_i T}$  is the submatrix of  $A$  that preserves the  $G$  rows corresponding to the group  $\mathcal{G}_i$  and the  $|T|$  columns corresponding to the sparsity set  $T$ . Given the set  $T$  defining the sparse support, the penalty factor  $\gamma(A, T, \mathcal{G})$  is a measure of similarity among the rows of  $A_{\mathcal{G}_i T}$  for each  $i$ . For example, if the rows of  $A_{\mathcal{G}_i T}$  are equal for some  $i$ , we will have  $\gamma(A, T, \mathcal{G}) = G$ ; in contrast, if all rows of  $A_{\mathcal{G}_i T}$  are mutually orthogonal for each  $i$ , then we will have  $\gamma(A, T, \mathcal{G}) = \sqrt{G}$ .

#### 3.2 Recovery Guarantees

We now provide requirements on the number of measurements needed for successful recovery of the sparse signal  $\underline{x}$  when the subset  $\Omega$  of the measurement domain is built in a structured way. The following theorem is proven in the appendix.

**Theorem 2.** *Let  $A$  be an  $N \times N$  orthogonal matrix ( $A^H A = I$ ) with coherence  $\mu(A)$ . Fix an arbitrary subset  $T$  of the signal domain. Choose a subset  $\Omega$  of*

the measurement domain of size  $|\Omega| = M$  as the union of  $M/G$  groups from  $\mathcal{G} = \{\mathcal{G}_1, \dots, \mathcal{G}_{M/G}\}$  and a sign sequence  $z$  on  $T$ , both uniformly at random over all possible choices. Suppose that

$$M \geq \gamma(A, T, \mathcal{G}) \cdot \text{Const} \cdot \mu^3(A) N^{3/2} |T| \log(N/\delta). \quad (4)$$

Then with probability exceeding  $1 - \delta$ , every signal  $c_0$  supported on  $T$  with signs matching  $z$  can be recovered from  $y = A_{\Omega c_0}$  by solving the linear program (2), for any  $\delta > 0$ .

The theorem shows that for perfectly incoherent measurement and sparsity bases,  $\gamma(A, T, \mathcal{G})$  provides a multiplicative penalty on the number of measurements necessary for successful signal recovery due to the grouped structure of the incoherent measurement selection. Note that for a group size  $G = 1$  and for perfectly incoherent pair of bases  $V$  and  $U$  our result coincides with Theorem 1 as it is equivalent to drawing elements of  $\Omega$  uniformly at random. Additionally, when the rows of some  $A_{\mathcal{G}_i T}$  are identical, the additional measurements bundled within the group provide no additional information, and we must scale the number of measurements by  $G$  in the worst case. This observation agrees with the penalty value  $\gamma(A, T, \mathcal{G}) = G$  for such a matrix  $A_T$  and group structure  $\mathcal{G}$ .

### 3.3 Calculation of the Performance Metric

For a fixed sparsity set  $T$ , we can obtain lower and upper bounds on the value of  $\gamma(A, T, \mathcal{G})$  by leveraging the Pietsch Factorization theorem [6].

**Theorem 3.** *Each matrix  $B$  can be factored as  $B = FD$  where  $D$  is a nonnegative, diagonal matrix with  $\text{trace}(D^2) = 1$  and  $\|B\|_{\infty \rightarrow 2} \leq \|F\|_2 \leq K_p \|B\|_{\infty \rightarrow 2}$ , where  $K_p$  is a constant equal to  $\sqrt{\frac{\pi}{2}} \approx 1.25$  for the real field and  $\sqrt{\frac{4}{\pi}} \approx 1.13$  for the complex field.*

Since  $\|M\|_{2 \rightarrow 1} = \|M^H\|_{\infty \rightarrow 2}$ , thanks to the duality of the operator norms, we can find bounds on  $\gamma$  by performing Pietsch factorization of the matrices  $(\overline{A_{\mathcal{G}_i T}})^H = F_i D_i$ , for  $i = 1, \dots, N/G$ , where  $D_i$  is a nonnegative diagonal matrix with  $\text{trace}(D_i^2) = 1$ . The value of  $\gamma(A, T, \mathcal{G})$  can then be bounded by

$$\frac{1}{K_p} \max_i \|F_i\|_2 \leq \gamma(A, T, \mathcal{G}) \leq \max_i \|F_i\|_2, \quad (5)$$

The Pietsch factorization of matrix  $B$  can be performed by solving a semidefinite program [6]. As will be pointed out in the simulations in the next section, the looseness of the bound (5) can limit the utility of  $\gamma$ .

## 4 Simulations

In this section, we present simulation results for CS recovery from grouped incoherent measurements for signals with sparse Fourier domain representations.

Consider a discrete signal  $s$  of length  $N = 1100$  and sparsity  $|T| = 5\% \cdot N$ , sparse in the frequency domain, generated as a product of an orthonormal Fourier basis of size  $N \times N$  and a sparse coefficient vector  $\underline{c}$  with values of non-zero entries distributed uniformly:  $\sim \mathcal{U}(-1, 1)$ . Consider two different configurations for the grouped incoherent measurements:

$\mathcal{G}^1$ : 100 groups of size 11 were constructed such that the first sample of each of the groups was chosen out of the first 100 samples of  $s$ :  $\{s[1], \dots, s[n]\}$ , and the remaining 10 samples for each group were shifted with respect to the first sample by multiples of 100. More specifically,  $\mathcal{G}_i^1 = \{i, i + 100, i + 200, \dots, i + 1000\}$ . This configuration appears in the interference-robust compressive wide-band receiver application. The first sample corresponds to a random zero-crossing of a modulated interferer. Additional samples correspond to subsequent zero-crossings of the interferer's carrier.

$\mathcal{G}^2$ : 100 groups of size 11 were constructed such that each group contained 11 consecutive, adjacent samples. More specifically,  $\mathcal{G}_i^2 = \{s[i + (i-1) \cdot 11] : s[i \cdot 11]\}$ . Such configuration assumes that the samples are taken in sequential bursts.

Figure 2 shows the relation between the penalty factor  $\gamma(A, T, \mathcal{G})$  from (3) and the ratio between the number  $M$  of samples required for successful recovery for the two described group structures and the number of samples  $M_0$  required for successful recovery for random sampling. The values shown are the minimal number of measurements needed to obtain normalized recovery error  $NRE = \|s - \hat{s}\|/\|s\| < 0.001$  for 99 out of 100 draws of the measurement groups (uniformly at random) and the values of the Fourier coefficients (from  $\mathcal{U}[-1, 1]$ ). Each point of the scatter plots corresponds to a fixed signal support. We consider three different classes of signal supports: for the first two types, the positions of the non-zero Fourier coefficients are chosen uniformly at random within a sub-band built out of two and four 5%-wide channels, respectively, positioned uniformly at random within the entire frequency band; we then compare their performance against the baseline of signals with unrestricted sparse supports. Figure 2 shows that for the first two classes  $\gamma$  was a good performance indicator; in contrast, for the last class the values of  $\gamma$  misleadingly suggest that both group structures perform equally well. Additionally, for a set of image processing experiments we performed but omit due to space constraints,  $\gamma$  was not a useful indicator of performance for group structures of practical interest. This is indicative of the potential looseness of the bound provided by Theorem 2. We believe that such looseness is a characteristic of guarantees that rely on worst-case metrics, such as the coherence  $\mu(A)$  and our grouped metric  $\gamma(A, T, G)$ , and is compounded by the looseness in the estimate of  $\gamma(A, T, G)$  obtained via Theorem 3 (of up to 21%).

## 5 Conclusions and Future Work

In this work, we have presented an analytically derived multiplicative penalty on the number of measurements needed for CS recovery when the measurements exhibit grouped structure instead of the usual independently drawn measure-

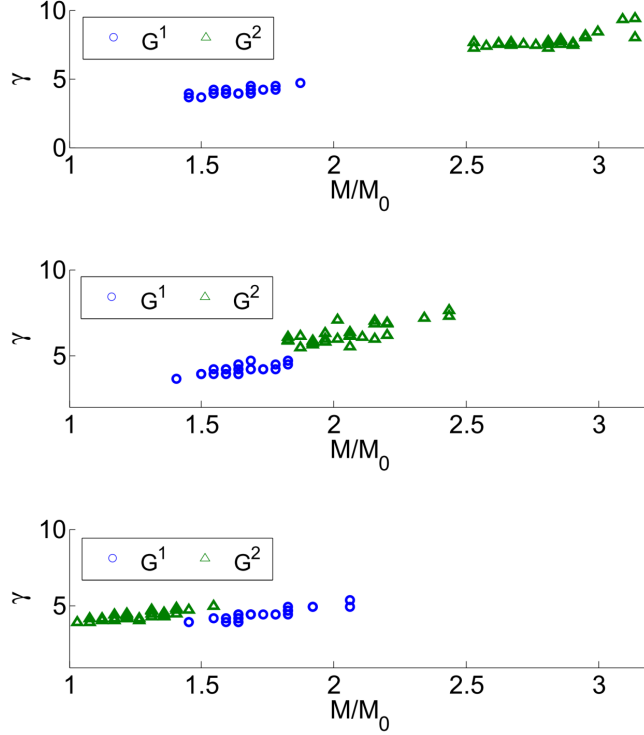


Figure 2:  $\gamma$  vs.  $M/M_0$  for group structures  $\mathcal{G}^1$  and  $\mathcal{G}^2$  when Fourier coefficients of 5% sparse signal  $s$  were concentrated within: (top) a sub-band built out of two 5%-wide channels; (middle) a sub-band built out of four 5%-wide channels; and (bottom) the entire band.

ment assumption taken by most existing CS literature. Such grouped sampling is of large practical interest as full randomization of measurements is difficult to achieve in many compressive sensing acquisition systems. A notable limitation of the introduced penalty factor  $\gamma$  is that it is dependent on the signal support. We expect further work to focus on penalty metrics that are independent of the support of the signal being measured, and to expand the guarantees provided to more applicable approximately sparse signals and to noisy measurement schemes.

## A Proof of Theorem 2

In this appendix, we will prove the result of Theorem 2 with a small modification on the distribution of the submatrices: Instead of a uniform distribution among all subsets  $\Omega$  containing  $M/G$  out of the  $N/G$  available groups, we pose



an independent Bernoulli selection for each group submatrix  $\mathcal{G}_i$ ,  $i = 1, \dots, N/G$ , belonging in  $\Omega$  with selection probability  $\delta_i = M/N$ . This independent model results in the expected number of selected groups being equal to  $M/G$ . Furthermore, one can show that since the probability of failure is a non-increasing function of the size  $M$  of the set  $\Omega$ , the probability of failure under the uniform distribution used in Theorem 2 is upper-bounded by a constant times the probability of failure under the independent selection model used in the proof (a property dubbed *poissonization* in [7]). Thus, the effect of the conversion of the subgroup selection model is a constant multiplicative factor in the required number of measurements, which is accounted for by the constant in (1).

Following the argument of [1], one can show that the signal  $\mathbf{c}_0$  is the unique solution to (2) if and only if there exists a dual vector  $\pi \in \mathbb{R}^N$  that has following properties:

- $\pi$  is in the row space of  $A_\Omega$ ,
- $\pi(t) = \text{sign}\{\mathbf{c}_0(t)\}$  for  $t \in T$ ,
- $|\pi(t)| < 1$  for  $t \in T^c$ .

As in [1], we consider the candidate

$$\pi = A_\Omega^H A_{\Omega T} (A_{\Omega T}^H A_{\Omega T})^{-1} z_0, \quad (6)$$

where  $z_0$  is a  $|T|$ -dimensional vector whose entries are the signs of  $\mathbf{c}_0$  on  $T$ . To prove Theorem 2 we need to show that under its hypothesis: (i)  $A_{\Omega T}^H A_{\Omega T}$  is invertible and (ii)  $|\pi(t)| < 1$  for  $t \in T^c$ . We begin by showing that  $A_{\Omega T}^H A_{\Omega T}$  is invertible with high probability given the requirement (4) on  $M$ . In Theorem 4, we claim that if  $M$  is large enough then, on average, the matrix  $A_{\Omega T}^H A_{\Omega T}$  does not deviate much from  $\frac{M}{N}I$ , where  $I$  is the identity matrix.

**Theorem 4.** *Fix an arbitrary subset  $T$  of the signal domain. Define  $N/G$  index groups  $\mathcal{G} = \{\mathcal{G}_1, \dots, \mathcal{G}_{N/G}\}$  of the measurement domain, each of size  $G$ , and draw each group independently at random into a set  $\Omega$  with probability  $M/N$ . If*

$$M \geq \frac{28}{3} \cdot \gamma(A, T, \mathcal{G}) \cdot N \cdot \mu^2(A) \cdot |T| \log\left(\frac{|T|}{\delta}\right), \quad (7)$$

*with  $\gamma(A, T, \mathcal{G})$  introduced in (3), then*

$$P\left(\left\|\frac{N}{M} A_{\Omega T}^H A_{\Omega T} - I\right\| \geq \frac{1}{2}\right) < \delta, \quad (8)$$

*where  $\|\cdot\|$  denotes the spectral norm.*

*Proof.* Denote

$$Y := \frac{N}{M} A_{\Omega T}^H A_{\Omega T} - I = \frac{N}{M} \sum_{i=1}^{N/G} \delta_i A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} - I, \quad (9)$$

where  $\delta_i$  is a Bernoulli random variable with  $P(\delta_i = 1) = \frac{M}{N}$ . Because  $A^H A = I$ , we have

$$\sum_{i=1}^{N/G} A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} = I, \quad (10)$$

and so we can write

$$Y = \frac{N}{M} \sum_{i=1}^{N/G} \left( \delta_i - \frac{M}{N} \right) A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} =: \sum_{i=1}^{N/G} Y_i. \quad (11)$$

We will use Theorem 1.4 from [8], which we include below for completeness.

**Theorem 5.** *Consider a finite sequence  $\{Y_i\}$  of independent self-adjoint random matrices with dimension  $d$ . Assume that each matrix  $Y_i$  satisfies  $\mathbb{E}\{Y_i\} = 0$  and  $\|Y_i\| \leq B$  almost surely. Then, for all  $t \geq 0$ ,*

$$P \left\{ \left\| \sum_i Y_i \right\| > t \right\} \leq d \cdot \exp \left( \frac{-t^2/2}{\sigma^2 + Bt/3} \right) \quad \text{where} \quad \sigma^2 = \left\| \sum_i \mathbb{E} Y_i^2 \right\|.$$

For our case,

$$Y_i = \left( \delta_i - \frac{M}{N} \right) A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} \frac{N}{M} \quad (12)$$

and  $E(Y_i) = 0$ . We find a bound  $B$  on  $\|Y_i\|$ :

$$\begin{aligned} \|Y_i\| &= \sup_{f_1, f_2} |\langle f_1, Y_i f_2 \rangle| \leq \sup_{f_1, f_2} \left| \left\langle f_1, \frac{N}{M} \sum_{l \in \mathcal{G}_i} a^l \otimes a^l f_2 \right\rangle \right| \\ &\leq \frac{N}{M} \sup_{f_1, f_2} \sum_{l \in \mathcal{G}_i} |\langle f_1, a^l \otimes a^l f_2 \rangle| = \frac{N}{M} \sup_{f_1, f_2} \sum_{l \in \mathcal{G}_i} |\langle f_1, a^l \rangle \langle a^l, f_2 \rangle| \\ &\leq \frac{N}{M} \|a^l\|^2 \sup_{f_2} \sum_{l \in \mathcal{G}_i} \frac{|\langle a^l, f_2 \rangle|}{\|a^l\|} \leq \frac{N}{M} \mu^2(A) |T| \gamma =: B, \end{aligned} \quad (13)$$

where the supremum is over unit-norm vectors  $f_1$  and  $f_2$ ,  $a^l$  is the  $l^{th}$  row of the matrix  $A_T$ , and  $\gamma$  is defined in (3).

Next, we calculate  $\sigma^2$  from Theorem 5 as

$$\sigma^2 = \left\| \sum_{i=1}^{N/G} \mathbb{E}(Y_i^2) \right\| = \text{var}\{\delta_i\} \frac{N^2}{M^2} \left\| \sum_{i=1}^{N/G} (A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T})^2 \right\|. \quad (14)$$

Since  $A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T}$  is a Hermitian matrix, its eigendecomposition is  $A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} = \Omega_i \Lambda_i \Omega_i^H$ , where  $\Omega_i$  is a matrix whose columns are the orthonormal eigenvectors  $\omega_{ij}$ ,  $j = 1, \dots, |T|$ , of the matrix  $A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T}$ , and  $\Lambda_i$  is a diagonal matrix containing the eigenvalues  $\{\lambda_{ij}\}_{j=1}^{|T|}$  of the matrix  $A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T}$ . Thus, we can write

$$A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} = \Omega_i \Lambda_i \Omega_i^H \Omega_i \Lambda_i \Omega_i^H = \Omega_i \Lambda_i^2 \Omega_i^H = \sum_{j=1, \dots, |T|} \lambda_{ij}^2 \omega_{ij} \omega_{ij}^H, \quad (15)$$

and so

$$\left\| \sum_{i=1}^{N/G} A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} \right\| = \left\| \sum_{i=1}^{N/G} \sum_{j=1, \dots, |T|} \lambda_{ij}^2 \omega_{ij} \omega_{ij}^H \right\|. \quad (16)$$

The right side of (16) is a weighted double sum of positive semidefinite matrices. The spectral norm of such a sum increases monotonically with the value of each of the weighting coefficients. Therefore, we can upper-bound (16) by replacing  $\lambda_{ij}$  with  $\max_{i=1, \dots, \frac{N}{G}} \max_{j=1, \dots, |T|} \lambda_{ij}$  and taking it out of the operator norm:

$$\begin{aligned} \left\| \sum_{i=1}^{N/G} A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} \right\| &\leq \max_{i=1, \dots, \frac{N}{G}} \max_{j=1, \dots, |T|} \lambda_{ij} \left\| \sum_{i=1}^{N/G} \sum_{j=1, \dots, |T|} \lambda_{ij} \omega_{ij} \omega_{ij}^H \right\| \\ &= \max_{i=1, \dots, \frac{N}{G}} \|A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T}\| \cdot \left\| \sum_{i=1}^{N/G} A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} \right\|. \end{aligned} \quad (17)$$

With (17) we can bound (14) by

$$\begin{aligned} \sigma^2 &\leq \text{var}\{\delta_i\} \frac{N^2}{M^2} \max_{i=1, \dots, \frac{N}{G}} \|A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T}\| \cdot \left\| \sum_{i=1}^{N/G} A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T} \right\| \\ &= \text{var}\{\delta_i\} \frac{N^2}{M^2} \max_{i=1, \dots, \frac{N}{G}} \|A_{\mathcal{G}_i T}^H A_{\mathcal{G}_i T}\| \cdot \|A_T^H A_T\| \\ &= \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N^2}{M^2} \max_{i=1, \dots, \frac{N}{G}} \sup_{\|f_1\|=\|f_2\|=1} |\langle f_1, \sum_{l \in \mathcal{G}_i} a^l \otimes a^l f_2 \rangle| \\ &\leq \frac{N}{M} \max_{i=1, \dots, \frac{N}{G}} \sup_{\|f_1\|=\|f_2\|=1} \sum_{l \in \mathcal{G}_i} |\langle f_1, a^l \otimes a^l f_2 \rangle| \\ &= \frac{N}{M} \max_{i=1, \dots, \frac{N}{G}} \sup_{\|f_1\|=\|f_2\|=1} \sum_{l \in \mathcal{G}_i} |\langle f_1, a^l \rangle| \cdot |\langle a^l, f_2 \rangle| \\ &\leq \frac{N}{M} \|a^l\|^2 \max_{i=1, \dots, \frac{N}{G}} \sup_{\|f_1\|=1} \sum_{l \in \mathcal{G}_i} \frac{|\langle f_1, a^l \rangle|}{\|a^l\|} \\ &\leq \frac{N}{M} \mu^2(A) |T| \max_{i=1, \dots, \frac{N}{G}} \sup_{\|f_1\|=1} \sum_{l \in \mathcal{G}_i} \frac{|\langle f_1, a^l \rangle|}{\|a^l\|} \\ &= \frac{N}{M} \mu^2(A) |T| \gamma = B. \end{aligned} \quad (18)$$

We put together (13), (18) and Theorem 5 to write

$$P(\|Y\| \geq 1/2) \leq |T| \cdot \exp \left\{ \frac{-1/8}{7/6 \cdot \frac{N}{M} \mu^2(A) |T| \gamma} \right\}, \quad (19)$$

which proves Theorem 4.  $\square$

Theorem 4 shows that if  $M$  is large enough, then  $A_{\Omega T}^H A_{\Omega T}$  is invertible with high probability. We continue by proving that  $|\pi(t)| < 1$  for  $t \in T^c$ . Following the techniques in [1], we need to prove four lemmas:

**Lemma 1.** Denote by  $v^0$  a row of the matrix  $A_{\Omega}^H A_{\Omega T}$  indexed by  $t_0 \in T^c$ . Then

$$E\|v^0\|^2 < \frac{M}{\sqrt{N}} \mu^3(A) |T| \gamma. \quad (20)$$

*Proof.* One can express  $v^0$  as

$$v^0 = \sum_{i=1}^{N/G} \delta_i \sum_{l \in \mathcal{G}_i} A(l, t_0) a^l. \quad (21)$$

Now due to the orthogonality of the columns of the matrix  $A$ ,

$$\sum_{i=1}^{N/G} \sum_{l \in \mathcal{G}_i} A(l, t_0) A(l, t) = 0, \quad (22)$$

and we can write

$$v^0 = \sum_{i=1}^{N/G} (\delta_i - E(\delta_i)) \sum_{l \in \mathcal{G}_i} A(l, t_0) a^l = \sum_{i=1}^{N/G} Y_i, \quad (23)$$

with

$$Y_i := \left( \delta_i - \frac{M}{N} \right) \sum_{l \in \mathcal{G}_i} A(l, t_0) a^l. \quad (24)$$

We see that  $E(Y_i) = 0$  and we can write

$$\begin{aligned} E\|v^0\|^2 &= E \left\| \sum_{i=1}^{N/G} Y_i \right\|^2 = E \left\langle \sum_{i=1}^{N/G} Y_i, \sum_{i'=1}^{N/G} Y_i \right\rangle \\ &= E \sum_{i=1}^{N/G} \langle Y_i, Y_i \rangle + E \sum_{i \neq i'}^{N/G} \langle Y_i, Y_{i'} \rangle = \sum_{i=1}^{N/G} E \langle Y_i, Y_i \rangle. \end{aligned} \quad (25)$$

Each element of the sum above can be bounded by

$$\begin{aligned} E \langle Y_i, Y_i \rangle &= E \left\langle \left( \delta_i - \frac{M}{N} \right) \sum_{l \in \mathcal{G}_i} A(l, t_0) a^l, \left( \delta_i - \frac{M}{N} \right) \sum_{l' \in \mathcal{G}_i} A(l', t_0) a^{l'} \right\rangle \\ &= \text{var}\{\delta_i\} \left\langle \sum_{l \in \mathcal{G}_i} A(l, t_0) a^l, \sum_{l' \in \mathcal{G}_i} A(l', t_0) a^{l'} \right\rangle \\ &= \frac{M}{N} \left( 1 - \frac{M}{N} \right) \sum_{l \in \mathcal{G}_i} A(l, t_0) \sum_{l' \in \mathcal{G}_i} \overline{A(l', t_0)} \langle a^l, a^{l'} \rangle \end{aligned}$$

$$\begin{aligned}
&\leq \frac{M}{N} \sum_{l \in \mathcal{G}_i} |A(l, t_0)| \sum_{l' \in \mathcal{G}_i} |A(l', t_0)| \cdot \left| \langle a^l, a^{l'} \rangle \right| \\
&\leq \mu(A) \frac{M}{N} \sum_{l \in \mathcal{G}_i} |A(l, t_0)| \sum_{l' \in \mathcal{G}_i} \left| \langle a^l, a^{l'} \rangle \right| \\
&\leq \mu(A) \frac{M}{N} \sum_{l \in \mathcal{G}_i} |A(l, t_0)| \sum_{l' \in \mathcal{G}_i} \frac{\left| \langle a^l, \frac{a^{l'}}{\|a^{l'}\|} \rangle \right|}{\|a^l\|} \cdot \|a^l\| \cdot \|a^{l'}\| \\
&\leq \mu^3(A) |T| \frac{M}{N} \sum_{l \in \mathcal{G}_i} |A(l, t_0)| \sum_{l' \in \mathcal{G}_i} \frac{\left| \langle a^l, \frac{a^{l'}}{\|a^{l'}\|} \rangle \right|}{\|a^l\|} \\
&\leq \mu^3(A) |T| \frac{M}{N} \gamma \sum_{l \in \mathcal{G}_i} |A(l, t_0)|.
\end{aligned} \tag{26}$$

Putting together (25) and (26), we get

$$\begin{aligned}
E \|v^0\|^2 &\leq \frac{M}{N} \mu^3(A) |T| \gamma \sum_{i=1}^{N/G} \sum_{l' \in \mathcal{G}_i} |A(l', t_0)| \\
&= \frac{M}{N} \mu^3(A) |T| \gamma \sum_{l=1}^N |A(l, t_0)| = \frac{M}{N} \mu^3(A) |T| \gamma \|A(:, t_0)\|_1.
\end{aligned} \tag{27}$$

Now since  $A^H A = I$ , we have  $\|A(:, t_0)\|_2 = 1$ , and since for any vector  $h$  of length  $N$  we have  $\|h\|_1 \leq \sqrt{N} \|h\|_2$ , it follows that

$$E \|v^0\|^2 \leq \frac{M}{\sqrt{N}} \mu^3(A) |T| \gamma, \tag{28}$$

which proves Lemma 1.  $\square$

**Lemma 2.** *Define*

$$\bar{\sigma}^2 := \gamma \mu^2(A) \frac{M}{N} \cdot \max \left\{ 1, \sqrt{\gamma} \mu^{3/2}(A) N^{3/4} |T| / \sqrt{M} \right\}. \tag{29}$$

For  $0 < a \leq \frac{\sqrt{M}}{\mu(A) \sqrt{N \gamma |T|}}$  if  $\frac{\sqrt{\gamma} \mu^{3/2}(A) N^{3/4} |T|}{\sqrt{M}} < 1$  and  $0 < a \leq \left( \frac{M}{\gamma \mu(A) \sqrt{N}} \right)^{1/4}$  if  $\frac{\sqrt{\gamma} \mu^{3/2}(A) N^{3/4} |T|}{\sqrt{M}} \geq 1$ , we have

$$P \left( \|v^0\| > \mu^{3/2}(A) N^{-1/4} \sqrt{\gamma M |T|} + a \bar{\sigma} \right) < 3e^{-\kappa a^2} \tag{30}$$

for some positive constant  $\kappa$ .

*Proof.* By definition,

$$\|v^0\| = \sup_{\|f\|=1} \langle v^0, f \rangle = \sup_{\|f\|=1} \sum_{i=1}^{N/G} \langle Y_i, f \rangle, \tag{31}$$

with  $v^0$  from (23) and  $Y_i$  from (24). For completeness, we reproduce below Theorem 3.2 from [1], which we use to prove Lemma 2.

**Theorem 6.** *Let  $Y_1, \dots, Y_N$  be a sequence of independent random variables taking values in a Banach space and let  $Z$  be the supremum  $Z = \sup_{f \in \mathcal{F}} \sum_{n=1}^N f(Y_n)$ , where  $\mathcal{F}$  is a countable family of real-valued functions. Assume that  $|f(Y)| < B$  for every  $f \in \mathcal{F}$  and all  $Y$ , and  $\mathbb{E}f(Y_i) = 0$  for every  $f \in \mathcal{F}$  and  $i = 1, \dots, N$ . Then, for all  $t \geq 0$ ,*

$$P(|Z - \mathbb{E}Z| > t) \leq 3 \exp \left( \frac{-t}{KB} \log \left( 1 + \frac{Bt}{\sigma^2 + B\mathbb{E}\bar{Z}} \right) \right), \quad (32)$$

where  $\sigma^2 = \sup_{f \in \mathcal{F}} \sum_{i=1}^N \mathbb{E}f^2(Y_i)$ ,  $\bar{Z} = \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^N f(Y_i) \right|$ , and  $K$  is a numerical constant.

Denote the mapping  $\langle Y_i, f \rangle$  for a fixed unit vector  $f$  as  $f(Y_i)$ , so that  $\bar{Z} = \sup_{\|f\|=1} \sum_{i=1}^{N/G} f(Y_i) = \|v^0\|$ . We have  $\mathbb{E}\{f(Y_i)\} = 0$ , and

$$\begin{aligned} |f(Y_i)| &= \left| \left\langle \left( \delta_i - \frac{M}{N} \right) \sum_{l \in \mathcal{G}_i} A(l, t_0) a^l, f \right\rangle \right| = \left| \left( \delta_i - \frac{M}{N} \right) \sum_{l \in \mathcal{G}_i} A(l, t_0) \langle a^l, f \rangle \right| \\ &< \left| \sum_{l \in \mathcal{G}_i} A(l, t_0) \langle a^l, f \rangle \right| \leq \max_{l \in \mathcal{G}_i} |A(l, t_0)| \cdot \max_{l \in \mathcal{G}_i} \|a^l\| \cdot \sum_{l \in \mathcal{G}_i} \frac{|\langle a^l, f \rangle|}{\|a^l\|} \\ &\leq \gamma \cdot \mu^2(A) \cdot \sqrt{|T|} =: B. \end{aligned} \quad (33)$$

Now we find a bound on  $\mathbb{E}\{f^2(Y_i)\}$ :

$$\begin{aligned} \mathbb{E}\{f^2(Y_i)\} &= \mathbb{E} \left\{ \left| \left\langle \left( \delta_i - \frac{M}{N} \right) \sum_{l \in \mathcal{G}_i} A(l, t_0) a^l, f \right\rangle \right|^2 \right\} \\ &= \text{var}\{\delta_i\} \cdot \left| \sum_{l \in \mathcal{G}_i} A(l, t_0) \langle a^l, f \rangle \right|^2 \\ &= \frac{M}{N} \left( 1 - \frac{M}{N} \right) \cdot \left| \sum_{l \in \mathcal{G}_i} A(l, t_0) \langle a^l, f \rangle \right|^2 \\ &\leq \frac{M}{N} \mu^2(A) \sum_{l \in \mathcal{G}_i} |\langle a^l, f \rangle|^2, \end{aligned}$$

and so

$$\sum_{i=1}^{N/G} \mathbb{E}\{f^2(Y_i)\} \leq \mu^2(A) \frac{M}{N} \sum_{i=1}^{N/G} \sum_{l \in \mathcal{G}_i} |\langle a^l, f \rangle|^2.$$

We know that  $\sum_{i=1}^{N/G} \sum_{l \in \mathcal{G}_i} |\langle a^l, f \rangle|^2 = 1$ ; therefore,

$$\sum_{i=1}^{N/G} \mathbb{E}\{f^2(Y_i)\} \leq \frac{M}{N} \mu^2(A) =: \sigma^2. \quad (34)$$

Plugging (33) and (34) in Lemma 1, we have

$$\mathbb{E}\{\bar{Z}\} = \mathbb{E}\{\|v^0\|\} \leq \mu^{3/2}(A)\sqrt{|T|}\sqrt{\gamma}\frac{\sqrt{M}}{N^{1/4}}. \quad (35)$$

Assume that  $\frac{\sqrt{\gamma}|T|\mu^{3/2}(A)N^{3/4}}{\sqrt{M}} < 1$  and  $0 < a \leq \frac{\sqrt{M}}{\sqrt{\gamma}\mu(A)\sqrt{N}\sqrt{|T|}}$ ; then with (29), we have

$$\bar{\sigma}^2 = \gamma\mu^2(A)\frac{M}{N} > B\mathbb{E}\{\bar{Z}\} = B\mathbb{E}\{\|v^0\|\} > B\mu^{3/2}(A)\sqrt{\gamma}\sqrt{|T|}\frac{\sqrt{M}}{N^{1/4}}, \quad (36)$$

and by writing  $t = a\bar{\sigma}$  we have

$$Bt = Ba\bar{\sigma} \leq \bar{\sigma}^2. \quad (37)$$

For  $\frac{\sqrt{\gamma}|T|\mu^{3/2}(A)N^{3/4}}{\sqrt{M}} > 1$  and  $0 < a \leq \left(\frac{M}{\gamma\mu(A)\sqrt{N}}\right)^{1/4}$ , with (29) we have

$$\begin{aligned} \bar{\sigma}^2 &= \gamma^{3/2}\mu^{7/2}(A)|T|\sqrt{M} = B\mathbb{E}\{\bar{Z}\} = B\mathbb{E}\{\|v^0\|\} \\ &= B\mu^{3/2}(A)\sqrt{\gamma}\sqrt{|T|}\frac{\sqrt{M}}{N^{1/4}} \end{aligned} \quad (38)$$

and so

$$Bt \leq \bar{\sigma}^2. \quad (39)$$

Putting together (37), (39), and Theorem 6, we can write

$$P\left(\|v^0\| > \mu^{3/2}(A)N^{-1/4}\sqrt{\gamma M|T|} + a\bar{\sigma}\right) < 3e^{-\kappa a^2}, \quad (40)$$

where  $\kappa$  is a numerical constant  $\kappa = \frac{\log(1.5)}{K}$  and  $K$  comes from Theorem 6. This completes the proof of Lemma 2.  $\square$

**Lemma 3.** Let  $w^0 = (A_{\Omega T}^H A_{\Omega T})^{-1}v^0$ . With the notations and assumptions of Lemma 2 we have:

$$P\left(\sup_{t_0 \in T^c} \|w^0\| \geq 2N^{3/4}\mu^{3/2}\sqrt{\frac{\gamma|T|}{M}} + \frac{2Na\bar{\sigma}}{M}\right) \leq ne^{-\gamma a^2} + P\left(\|A_{\Omega T}^H A_{\Omega T}\| \leq \frac{M}{2N}\right).$$

*Proof.* Denote the events

$$E_1 : \left\{ \|A_{\Omega T}^H A_{\Omega T}\| \geq \frac{M}{2N} \right\}$$

and

$$E_2 : \left\{ \sup_{t_0 \in T^c} \|v^0\| \leq \mu^{3/2}(A)N^{-1/4}\sqrt{\gamma M|T|} + a\bar{\sigma} \right\}.$$

We can write

$$\begin{aligned} P\left(\sup_{t_0 \in T^c} \|w^0\| \geq 2N^{3/4}\mu^{3/2}\sqrt{\gamma|T|/M} + 2Na\bar{\sigma}/M\right) &\leq P(\overline{E_1 \cap E_2}) \\ &= P(\overline{E_1} \cup \overline{E_2}), \end{aligned} \quad (41)$$

and with Lemma 2 we have

$$\begin{aligned} P\left(\sup_{t_0 \in T^c} \|w^0\| \geq 2N^{3/4}\mu^{3/2}\sqrt{\frac{\gamma|T|}{M}} + \frac{2Na\bar{\sigma}}{M}\right) &\leq P\left(\|A_{\Omega T}^H A_{\Omega T}\| \leq \frac{M}{2N}\right) \\ &\quad + 3e^{-\kappa a^2}. \end{aligned} \quad (42)$$

□

Finally we will use Lemma 3.4 from [1]

**Lemma 4.** *Assume that  $z(t)$ ,  $t \in T$  is an i.i.d. sequence of symmetric Bernoulli random variables. For each  $\lambda > 0$ , we have*

$$P\left(\sup_{t \in T^c} |\pi(t)| > 1\right) \leq 2ne^{-1/2\lambda^2} + P\left(\sup_{t \in T^c} \|w^0\| > \lambda\right). \quad (43)$$

Now that all lemmas are in place, we are ready to prove Theorem 2. If we pick  $\lambda = 2N^{3/4}\mu^{3/2}\sqrt{\gamma|T|/M} + 2Na\bar{\sigma}/M$  in (43), from (42) and (43) we get

$$P\left(\sup_{t \in T^c} |\pi(t)| > 1\right) \leq 2Ne^{-1/2\lambda^2} + Ne^{-\kappa a^2} + P\left(\|A_{\Omega T}^H A_{\Omega T}\| \leq M/2N\right). \quad (44)$$

For the right hand side of (44) to be smaller than  $3\delta$  we need all three summands to be smaller than  $\delta$ . We now derive conditions on  $\delta$  that provide this guarantee. We start with the second summand: for it to be no bigger than  $\delta$  we can set  $a^2$  to be

$$a^2 = \kappa^{-1} \log(N/\delta). \quad (45)$$

For the first summand to be no bigger than  $\delta$ , we need

$$\frac{1}{\lambda^2} \geq 2 \log(2N/\delta). \quad (46)$$

If  $\frac{\sqrt{\gamma}\mu^{3/2}(A)N^{3/4}|T|}{\sqrt{M}} > 1$ , Lemma 2 requires

$$0 < a \leq \left(\frac{M}{\gamma\mu(A)\sqrt{N}}\right)^{1/4}. \quad (47)$$

Then with  $\bar{\sigma}^2$  from (29) we get

$$Na\bar{\sigma}/M \leq \mu^{3/2}N^{3/4}\sqrt{\gamma|T|/M}, \quad (48)$$



and so

$$\lambda \leq 4\mu^{3/2}(A)N^{3/4}\sqrt{\gamma|T|/M}. \quad (49)$$

Reorganizing terms, we obtain

$$\frac{1}{\lambda^2} \geq \frac{M}{16\mu^3(A)N^{3/2}\gamma|T|}. \quad (50)$$

From (45) and (47) we get the following bound on  $M$ :

$$M \geq \gamma\mu(A)\sqrt{N}\kappa^{-2}\log^2(N/\delta). \quad (51)$$

Suppose now that  $\frac{\sqrt{\gamma}\mu^{3/2}(A)N^{3/4}|T|}{\sqrt{M}} < 1$ . Then, with (29),  $\bar{\sigma}^2 = \gamma\mu^2(A)\frac{M}{N}$ . If  $\mu(A)\sqrt{N}|T| \geq a^2$ , then

$$Na\bar{\sigma}/M \leq \mu^{3/2}N^{-1/4}\sqrt{\gamma|T|/M}, \quad (52)$$

and

$$\lambda \leq 4\mu^{3/2}(A)N^{3/4}\sqrt{\gamma|T|/M}, \quad (53)$$

and thus

$$\frac{1}{\lambda^2} \geq \frac{M}{16\mu^3(A)N^{3/2}\gamma|T|}, \quad (54)$$

which matches the previous condition (50). On the other hand, if  $\mu(A)\sqrt{N}|T| \leq a^2$  then

$$Na\bar{\sigma}/M \geq \mu^{3/2}N^{-1/4}\sqrt{\gamma|T|/M}, \quad (55)$$

and

$$\lambda \leq 4Na\bar{\sigma}/M, \quad (56)$$

and thus, with  $\bar{\sigma}^2 = \gamma\mu^2(A)\frac{M}{N}$ ,

$$\frac{1}{\lambda^2} \geq \frac{M^2}{16N^2a^2\bar{\sigma}^2} = \frac{M}{16a^2\gamma\mu^2(A)N}. \quad (57)$$

And so with (54) and (57) we can write

$$\frac{M}{16\gamma\mu^2(A)N} \min\left(\frac{1}{\mu(A)N^{1/2}|T|}, \frac{1}{a^2}\right) \geq 2\log(2N/\delta), \quad (58)$$

$$M \geq 16\gamma\mu^2(A)N \max\left(\mu(A)N^{1/2}|T|, a^2\right) 2\log(2N/\delta), \quad (59)$$

which with (45) gives

$$M \geq \text{Const} \cdot \gamma\mu^2(A)N \max\left(\mu(A)N^{1/2}|T|, \log\left(\frac{N}{\delta}\right)\right) \log\left(\frac{N}{\delta}\right). \quad (60)$$

Due to Theorem 4, for the third summand to be smaller than  $\delta$ , we need

$$M \geq \frac{28}{3} \cdot \gamma \cdot N \cdot \mu^2(A) \cdot |T| \log\left(\frac{|T|}{\delta}\right). \quad (61)$$

Thus from (51), (60) and (61) we see that the overall requirement on  $M$  is:

$$M \geq \text{Const} \cdot \gamma(A, T, \mathcal{G}) \cdot \mu^3(A)N^{3/2}|T|\log(N/\delta), \quad (62)$$

which finishes the proof of Theorem 2.

## References

- [1] E. J. Candès and J. K. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse Problems*, vol. 23, no. 3, pp. 969–985, June 2007.
- [2] M. A. Davenport, P. T. Boufounos, and R. G. Baraniuk, “Compressive domain interference cancellation,” in *Workshop on Signal Proc. with Adaptive Sparse Structured Representations (SPARS)*, Saint Malo, France, Apr. 2009.
- [3] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing MRI,” *IEEE Signal Proc. Mag.*, vol. 25, no. 3, pp. 72–82, Mar. 2008.
- [4] M. F. Duarte and Y. C. Eldar, “Structured compressive sensing: From theory to applications,” *IEEE Trans. Signal Proc.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.
- [5] V. M. Patel, G. R. Easley, Jr. D. M. Healy, and R. Chellappa, “Compressed synthetic aperture radar,” *IEEE J. Selected Topics in Signal Proc.*, vol. 4, no. 2, pp. 244–254, Apr. 2010.
- [6] J. A. Tropp, “Column subset selection, matrix factorization, and eigenvalue optimization,” in *ACM-SIAM Symp. on Discrete Algorithms (SODA)*, New York, NY, Jan. 2009, pp. 978–986.
- [7] E. J. Candès, J. K. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, Feb. 2006.
- [8] J. A. Tropp, “User-friendly tail bounds for sums of random matrices,” *Foundations of Computational Mathematics*, August 2011, To be published.